



HAL
open science

Shortening trinucleotide repeats using highly specific endonucleases: a possible approach to gene therapy?

Guy-Franck Richard

► **To cite this version:**

Guy-Franck Richard. Shortening trinucleotide repeats using highly specific endonucleases: a possible approach to gene therapy?. *Trends in Genetics*, 2015, 31 (4), pp.177 - 186. 10.1016/j.tig.2015.02.003 . pasteur-01370705

HAL Id: pasteur-01370705

<https://pasteur.hal.science/pasteur-01370705>

Submitted on 23 Sep 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1
2
3
4
5 **SHORTENING TRINUCLEOTIDE REPEATS USING HIGHLY SPECIFIC**
6
7
8 **ENDONUCLEASES: A POSSIBLE APPROACH TO GENE THERAPY ?**
9

10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27 **Guy-Franck Richard**
28
29

30
31
32 Institut Pasteur, Department Genomes & Genetics, CNRS UMR3525, 25 rue du Dr Roux, 75015 Paris
33
34
35
36
37
38
39
40

41
42 Keywords: Tandem repeats; Meganuclease; ZFN; TALEN; CRISPR-Cas; double-strand break
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

ABSTRACT

Trinucleotide repeat expansions are involved in more than two dozens neurological and developmental disorders. Conventional therapeutic approaches aiming at regulating the expression level of affected genes, by using drugs, oligonucleotides and/or transgenes, obtained mitigated successes so far. An alternative approach relies on shortening repeats to non-pathological lengths, using highly specific nucleases. This review will focus on experiments using meganucleases, zinc-finger nucleases and TALENs to contract trinucleotide repeats, and will discuss the possibility of using CRISPR-Cas nucleases to the same end. Advantages and drawbacks of each nuclease will be debated, particularly in the light of their respective efficacy and specificity on repeated and structured DNA. Nuclease design and delivery in the perspective of future gene therapy will be discussed.

MORE THAN 20 YEARS OF TRINUCLEOTIDE REPEAT DISORDERS: WHAT DID WE LEARN ?

The first neurological disorders linked to trinucleotide repeats, fragile X syndrome ¹, myotonic dystrophy type 1 ^{2,3} and Huntington disease ⁴ were discovered more than twenty years ago, and found to be associated to an expansion of the trinucleotide repeat tract. Since then, the scientific community has been trying to unravel molecular mechanisms responsible for these large expansions, as well as finding cures for these dramatic disorders, with limited successes so far. With the subsequent discovery of disorders linked to the expansion of non-trinucleotide repeats, the "rule of three" was broken ⁵. At the present time, more than thirty neurological and developmental disorders have been shown to be induced by the expansion of a tri-, tetra-, penta-, hexa- or dodecanucleotide repeat ⁶⁻⁸. Expansion sizes vary from a few to thousands of repeat units, but in all cases, the pathology is always remarkably associated to the expansion of a repeat tract (Figure 1). Conversely, when an expanded CTG repeat was contracted below pathological length during transmission from a father affected by myotonic dystrophy type 1 to his daughter, no clinical sign of the pathology could be detected in the

1 daughter ⁹. Hence, at the turn of the century, rose the idea that artificially shortening an expanded
2 trinucleotide repeat, using a dedicated molecular tool, could be a way of reducing -or suppressing-
3 pathological symptoms linked to the expansion. Meanwhile, several groups studied molecular
4 mechanisms involved in trinucleotide repeat expansions and contractions, using model systems such as
5 *Escherichia coli*, *Saccharomyces cerevisiae*, transgenic mice and human cell cultures, as well as *in*
6 *vitro* biochemical approaches. Molecular mechanisms of instability as well as DNA-damage response
7 to chromosomal fragility induced by trinucleotide repeats, have been extensively described in previous
8 reviews ¹⁰⁻¹⁶, and will not be exposed here for the sake of conciseness, but I will rather focus on recent
9 developments in the field.

22 **Molecular mechanisms triggering trinucleotide repeat instability, from bacteria to mammals**

23 Basically any pathway involving *de novo* DNA synthesis of trinucleotide repeat-containing DNA may
24 generate expansions and contractions of the repeat tract. Replication ¹⁷⁻²³, double-strand break repair ²⁴,
25 ²⁵, base excision repair of oxydative damage ²⁶ and single-strand break repair ²⁷, all interfere with
26 propagation of normal size trinucleotide repeat tracts (Box 1). This holds true from bacteria to yeast
27 and transgenic mice or human cell lines, with subtle differences that will not be debated in the present
28 review. However, three mechanisms have recently drawn more attention: mismatch repair, helicase
29 unwinding and reversed replication forks, and their respective impact on trinucleotide repeat instability
30 will be more extensively discussed.

33 *Mismatch repair*

34 Mismatch repair is a sophisticated machinery made of different heterodimeric complexes in
35 eukaryotes, involved in detecting and fixing replication errors (reviewed in ²⁸. Early on, it was
36 described that Msh2 deficiency decreased the rate of expansions at a CAG/CTG trinucleotide repeat, in
37 transgenic mice ²⁹. This was further confirmed by subsequent studies in mouse ³⁰⁻³⁴ and *in vitro* models
38 ³⁵⁻³⁷. It was proposed that binding of Msh2 complexes to CAG/CTG hairpin structures ³⁸ would
39 stabilize the hairpin, inhibiting ³⁹ or not ⁴⁰ Msh2-Msh3 ATPase activity, and eventually leading to

1 repeat tract expansion. In yeast, a 2-3 fold increase in contractions ^{41, 42}, or a 10-fold decrease in
2 expansions ⁴³ were detected in *msh2Δ* mutants. These rather moderate effects may be explained by the
3 use of shorter repeats in yeast (25-50 triplets) as compared to mice experiments (100-300 triplets). It is
4 likely that trinucleotide repeat tract length needs to reach a certain threshold in order to form hairpins
5
6 efficiently recognized and bound by Msh2 complexes.
7
8
9

10 *Helicase unwinding*

11 It was shown that the yeast *SRS2* helicase selectively unwinds CAG/CTG trinucleotide repeats *in vitro*
12
13 ^{44, 45}, and that these repeats are destabilized in *srs2Δ* mutants ^{19, 46}. *SRS2* was also involved in
14
15 facilitating replication through CGG/CCG trinucleotide repeats in yeast ⁴⁷, indicating that the helicase
16
17 is structure- but not sequence-specific. Interestingly, the rate of GAA/TTC trinucleotide repeat
18
19 expansions is unchanged in a *srs2Δ* mutant ²⁰, suggesting that different secondary structures are
20
21 formed by CAG/CTG and GAA/TTC repeats *in vivo*. More recently, the human RTEL1 helicase,
22
23 involved in homologous recombination and telomere maintenance, was shown to complement
24
25 instability and fragility phenotypes of a *srs2Δ* mutant yeast strain, suggesting that this helicase could
26
27 play, in human cells, the same role as *SRS2* in yeast. When RTEL1 was knockdown, trinucleotide
28
29 repeat expansions were increased 5-7 fold in human cells, confirming results obtained in yeast ⁴⁸.
30
31 CAG/CTG repeat contractions increase in deletants of the *SGS1* gene, another yeast S phase helicase,
32
33 homologous to the Blm and Wrn human helicases ¹⁹, whereas GAA/TTC repeat expansions decrease 3
34
35 fold in a *sgs1Δ* mutant ²⁰. This suggests that this helicase has a lower specificity for secondary
36
37 structures, or that its effect is not to unwind structures, but to facilitate replication through them in
38
39 some other way.
40
41
42
43
44
45
46
47
48
49

50 *Reversed replication forks*

51 Replication fork reversal was first proposed to occur in *E. coli*, following replication fork arrest ⁴⁹.
52
53 Subsequently, unusual replication molecules that may be interpreted as reversed replication forks were
54
55 detected in slow replicating yeast cells in which the replication checkpoint was compromised ⁵⁰. Later
56
57 on, it was proposed that fork reversal may promote folding of trinucleotide repeat hairpins, that would
58
59
60
61
62
63
64
65

ultimately lead to repeat expansions ¹³. However, experiments in *E. coli* and yeast cells showed that, on the contrary, genetic factors that promote replication fork reversal tend to stabilize trinucleotide repeats ^{19,51}. It must be mentioned that formation of reversed forks during replication of a trinucleotide repeat tract has not been formally proven so far, but only inferred from unusual 2D gel structures ¹⁹. It is actually a real challenge to look at specific DNA molecules during replication of a unique locus, as compared to genome-wide structures, but interesting approaches are being developed that could address this question ⁵².

Formation of DNA and RNA secondary structures *in vitro* and *in vivo*

Early on, it was shown that trinucleotide repeats were able to fold into specific secondary structures *in vitro* ⁵³. It was subsequently proposed that these structures could be responsible for triggering the expansion process ⁵⁴. All CNG repeats studied fold into A-form DNA hairpins, in which non-canonical N-N pairs are stabilized by C-G pairs ⁵⁵. This holds true for RNA repeats ⁵⁶⁻⁵⁸. In addition, CGG repeats form tetrahelical structures ⁵⁹, as well as CCG repeats at neutral or slightly acidic pH ⁶⁰. Triple helices containing Watson-Crick as well as Hoogsteen base pairings are formed by GAA/TTC repeats ⁶¹, similarly to other regular homopurine-homopyrimidine tracts ⁶². At the present time, there is little direct evidence to support the presence of similar secondary structures in living cells, except for the existence of G-quadruplex formed by telomeric repeats ⁶³. However, it was recently shown that an antibody directed against slipped-stranded DNA structures was able to bind and immunoprecipitate expanded CAG/CTG repeats from myotonic dystrophy patient cells ⁶⁴, proving that slipped-strand structures exist *in vivo* within such repeats. It must also be mentioned that long trinucleotide repeats transiently stall replication forks in bacteria ⁶⁵ and yeast cells ^{20, 47, 66}, suggesting that secondary structures probably also impede replication fork progression *in vivo*. It cannot be formally excluded though, that proteins bound at the repeat locus are in fact -directly or indirectly- stalling the fork, rather than DNA secondary structures by themselves.

The role of chromatin.

1 One striking difference between yeast and mammalian models is the propensity of trinucleotide repeats
2
3 to expand in humans and in transgenic mice and to contract in yeast cells, even though the exact same
4
5 repeat tract is studied. One explanation could come from differences in chromatin structures between
6
7 yeast and mammalian cells. Actually, CAG/CTG trinucleotide repeat tracts exhibit a noncanonical
8
9 organization when embedded in a yeast chromosome. Hypersensitive sites to the Fnu4HI endonuclease
10
11 occur in a distinctly shorter spacing than the canonical 165 bp distance of yeast nucleosomes ⁶⁷. A
12
13 high-mobility group protein, Hmo1, was shown to bind specifically to the trinucleotide repeat tract,
14
15 modifying chromatin conformation at this locus. The human genome contains many HMG-Box
16
17 proteins, but it is unknown if one of them could play the same role as Hmo1 in yeast. More recently, it
18
19 was shown that both histone H4 acetylation and deacetylation were important to counter CAG/CTG
20
21 trinucleotide repeat expansions occurring during post-replication repair ⁶⁸, in yeast cells. However, the
22
23 possible interaction of H4 acetylation/deacetylation with modified chromatin in *hmo1Δ* mutant was not
24
25 tested. Interestingly, inhibiting histone deacetylase complexes HDAC3 and HDAC5 in human cells led
26
27 to a decrease of expansions ^{69, 70}. It is possible though that histones are not the only target of these
28
29 HDACs, and that other proteins need to be deacetylated in order to promote trinucleotide repeat
30
31 expansions. In addition to the role of deacetylation, it was also shown in human cells that decreasing
32
33 the level of CpG methylation by knocking down the maintenance methyltransferase Dnmt1, doubled
34
35 the frequency of CAG/CTG contractions, whereas expansions increased in the progeny of mice
36
37 carrying only one Dnmt1 functional allele ⁷¹. In conclusion, chromatin structure and associated
38
39 remodeling factors interfere strongly with trinucleotide repeat expansions, although the precise
40
41 mechanisms involved need to be clarified.
42
43
44
45
46
47
48
49
50
51
52
53

SHORTENING TRINUCLEOTIDE REPEATS USING HIGHLY SPECIFIC ENDONUCLEASES

54
55
56
57 With the discovery of the first meganuclease, I-Sce I ⁷², it was clear that highly specific nucleases
58
59 could be used for gene targeting in plant or mammalian cells, opening the way to genetic modified
60
61
62
63
64
65

organisms, as well as to gene therapy. At the present time, four different families of highly specific DNA endonucleases have been discovered and/or engineered : meganucleases, zinc-finger nucleases, TALENs and CRISPR-Cas9 nucleases. All of these have been used, with relative success, to engineer eukaryotic genomes. Extensive and thoughtful reviews are available for those who want to use these nucleases for genome engineering ^{73, 74}, but the present review will only focus on the scarce but promising experiments aiming at shortening trinucleotide repeat tracts using such highly specific endonucleases.

Meganucleases

In early experiments performed in haploid yeast cells, an I-*Sce* I recognition site inserted in the middle of a (CAG)₁₀ repeat was integrated in a yeast chromosome (Figure 2). A plasmid carrying regions of homology surrounding a CAG/CTG trinucleotide repeat was provided in *trans*, as a homologous template for recombination. When I-*Sce* I was induced, most of the yeast cells (67%) repaired the double-strand break (DSB) by annealing the repeats flanking the DSB, leading to final repair products of 4, 5 or 6 triplets long ⁴¹. In all other cases (33%), cells repaired the DSB using the donor plasmid as a template for gene conversion, leading to frequent contractions and expansions of the repeat tract ²⁵. It was later demonstrated that similar outcomes were observed during meiosis, when a DSB made by a meiotically induced I-*Sce* I gene was repaired using a trinucleotide repeat-containing homologue, although not in the same proportions as those observed during mitotic DSB-repair ²⁴. All these observations led to the conclusion that making a DSB within a trinucleotide repeat could be an efficient way of promoting contractions of the repeat tract (and sometimes expansions too). However, the development of new meganucleases in order to recognize repeated sequences like trinucleotide repeats has proven to be a tedious and hazardous task. Using the I-*Cre* I meganuclease as a backbone, Arnould and colleagues generated 42 new nucleases with DNA target specificities different from the original I-*Cre* I ⁷⁵. A comprehensive analysis of the efficacy of 37 of these new nucleases showed that their efficacy at targeted mutagenesis in cell culture was not correlated to their activity in transiently

transfected substrates, suggesting that chromatinian structures strongly modulate accession of the nuclease to its cognate site ⁷⁶. Given the probable structured nature of trinucleotide repeats *in vivo*, development of a meganuclease efficient at targeting such repeats seemed very unlikely.

Zinc-Finger Nucleases (ZFN)

Zinc-finger nucleases (ZFN) were originally built by fusion of a zinc-finger transcription activation domain to the cleavage domain of the *FokI* nuclease ⁷⁷. Each "finger" theoretically recognizes three nucleotides, and the *FokI* domain needs to dimerize in order to catalyze DNA cleavage. Two ZFN "arms" are used together spaced by a few nucleotides, so that binding of both arms on their cognate sequence will allow *FokI* dimerization and cutting. Two independent studies used ZFNs to induce DSBs into CAG/CTG trinucleotide repeat tracts. The first study used one arm recognizing the GCT triplet, whereas the second arm recognized the GCA sequence ⁷⁸. Expression of both arms in CHO cells carrying a trinucleotide repeat-containing reporter gene led to a 15-fold increase in repeat contraction frequency. However, expression of the GCT arm by itself already increased contractions 10 fold, suggesting efficient homodimerization of this arm on the repeat tract, probably by misrecognition of the GCA triplet by the GCT arm. Sequencing the repeat-containing locus after expression of the GCT arm showed frequent contractions of the repeat tract, in addition to deletions in one or both flanking regions (20% of the cases) or insertions of extra DNA sequences into the repeat tract (24% of the cases). In the second study, one arm recognizing the GCT triplet, the other arm recognizing the AGC sequence were designed ⁷⁹. When only one arm was expressed in HeLa cells, frequent repeat contractions were observed, that were interpreted by the authors as cleavage of CAG/CTG hairpins catalyzed by homodimerization of the arm. When both arms were expressed, contractions as well as expansions were recovered, intriguingly suggesting that DSB-repair mechanisms when homodimeric or heterodimeric ZFNs make the DSB, are different. It is also possible that the nature of the lesion itself was different (double-strand break *vs* single-strand nick). Both studies therefore concluded to a lack of specificity of ZFNs since only one arm was sufficient to

1 induce a DSB within a trinucleotide repeat tract. To alleviate this known default of ZFNs, new proteins
2 have been engineered, using four fingers instead of three, and a new *FokI* architecture. A *FokI* domain
3 variant was built, that can only function as an obligate heterodimer, reducing occurrences of genome-
4 wide off-target cleavage⁸⁰.
5
6
7
8
9

10 **Transcription-Activator Like Effector Nucleases (TALEN)**

11
12 Transcription-Activator Like Effectors (TALE) were discovered in the plant pathogen
13 *Xanthomonas*⁸¹. The DNA-binding domain of TALEs is made of tandemly repeated 34 amino-acid
14 motifs, identical to each other except for two residues located in the middle of each motif, called
15 repeat-variable diresidues (RVD). Each RVD recognizes, more or less specifically, A, T, C, or G^{82, 83}.
16
17 Addition of successive motifs allows to assemble TALEs that could theoretically recognize any DNA
18 sequence, including trinucleotide repeats. TALE Nucleases -or TALENs- were engineered by the
19 fusion of TALEs to the *FokI* nuclease domain⁸⁴. Like ZFNs, a TALEN is made of two arms, each arm
20 binding DNA sequences separated by a few nucleotides in order for the two *FokI* domains to dimerize
21 and catalyze DSB formation. A TALEN was built to be expressed in yeast cells containing a
22 CAG/CTG trinucleotide repeat from the human DM1 locus. The left arm was designed to bind the
23 junction between non-repeated and repeated DNA, while the right arm was designed to bind the
24 (CTG)₆ sequence. Expression of both arms in yeast cells led to trinucleotide repeat contractions in
25 more than 99% of the yeast colonies analyzed, whereas expression of the right arm along with a
26 mutated left arm had no effect⁸⁵. The genomes of seven yeast colonies in which the TALEN was
27 expressed were completely resequenced and no mutation other than the targeted repeat tract contraction
28 was detected. In particular, no other CAG/CTG repeat tract of the yeast genome showed any size
29 variation. No reciprocal translocation or large chromosomal rearrangement was found by pulse-field
30 gel electrophoresis. Based on these experiments in yeast, this TALEN seems to be very efficient, as
31 well as very specific, showing no evidence for off-target mutations. This needs, of course, to be now
32 confirmed in organisms exhibiting a more complex genome than *Saccharomyces cerevisiae*. New
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 TALEN architectures were recently developed, in order to build more compact single-molecule
2 nucleases, by replacing the *FokI* domain by a meganuclease domain. In one report, a TALE was linked
3 to the I-*Ani* I meganuclease. This new architecture combines TALE repeats and meganuclease
4 specificities, thus reducing potential off-target cleavages. It also allows to shorten the size of the TALE
5 region, making the fusion protein more compact. However, the TALE must target sequences that are
6 close to natural recognition sites of the I-*Ani* I nuclease ⁸⁶. Similar experiments were performed by
7 building a monomeric TALEN from the fusion of a TALE to the I-*Tev* I N-terminal region that
8 contributes to cleavage specificity ^{87, 88}. This fusion protein exhibited efficacies similar to the TALE-
9 *FokI* nuclease on the same targets, but was limited to cut degenerate I-*Tev* I cleavage sites (sequence:
10 CNNNGN).
11
12
13
14
15
16
17
18
19
20
21
22
23
24

25 **CRISPR-Cas9 nuclease: advantages and drawbacks**

26
27 At the same time these technical improvements were made to TALE nucleases, was solved the
28 function and mechanistics of CRISPR-Cas nucleases. Clustered regularly interspaced short
29 palindromic repeats (CRISPR) and their associated nucleases were discovered in bacteria and archaea,
30 providing these organisms with an adaptive immunity system against plasmids and phages ⁸⁹. Type II
31 CRISPR-Cas systems are made of a single nuclease (Cas9) and two small RNA molecules, the
32 CRISPR RNA (crRNA) containing the sequence to be targeted, and the trans-activating crRNA
33 (tracrRNA) that folds into a hairpin and pairs with the crRNA ⁹⁰. The system was subsequently
34 modified by fusion of both small RNAs into a single chimeric guide RNA (sgRNA or gRNA) filling
35 out both functions ⁹¹. The advantage of CRISPR-Cas9 over the aforementioned families of nucleases is
36 obviously the easiness of designing a new one, since only 20 nt homologous to the targeted region
37 need to be synthesized. This extreme simplicity and versatility (a nuclease-dead mutant will transform
38 it into a highly specific DNA-binding protein) explains in part the extraordinary success of CRISPR-
39 Cas9, at least in the litterature, with more than 400 publications mentioning the name CRISPR in 2014
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

(Box 2). However, they have not been used on trinucleotide repeat sequences yet, but two major obstacles may dramatically restrain their use to that end.

1
2
3 First, the nuclease strictly requires the presence of a protospacer adjacent motif (PAM) in the target
4 DNA ⁹². For the *Streptococcus pyogenes* Cas9 nuclease, which is the most widely used right now, the
5 PAM sequence is NGG and must therefore be present near the sequence to be targeted ⁹³. Given that
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
NGG is not present in many trinucleotide repeats (all poly-Gln and poly-Ala disorders, and most of non-coding repeats, see figure 1), it precludes its use, except if the nuclease is targeted at the junction between repeated and non-repeated DNA that could, by chance, contain a PAM sequence (Box 3). The second obstacle, that may be more dramatic for the use of CRISPR on trinucleotide repeat tracts, is the formation of secondary structures within the guide RNA itself, that could impede its recognition by Cas9 or recognition of the target sequence by the CRISPR-Cas9. As previously shown, all trinucleotide repeats (and other microsatellites) involved in human disorders are able to form secondary structures at the RNA level ⁵⁸. It is very likely that these structures will interfere with correct assembly and/or binding of the ribonucleoprotein itself, impeding its function. Design of a RNA guide that would cover the junction between repeated and non-repeated DNA may reduce the chance to form secondary structures, but is limited by the presence of a PAM sequence near the junction. Given these obstacles, trying to design a CRISPR-Cas9 aimed at targeting and shortening a trinucleotide repeat tract promises to be a tedious task, if at all possible.

COMPARING EFFICACY AND SPECIFICITY OF ENDONUCLEASES

Differences in protein architectures and implications for DNA recognition.

The four families of nucleases show very different architectures, as shown by their structures. LAGLIDAGD homing endonucleases contact DNA by a large number of interactions of two large $\beta\beta\beta\beta$ structures with the major groove, and of the N-terminal loop with the minor groove (Figure 3A). These many interactions explain I-Sce I (and more generally meganucleases) high specificity, but make the design of new nucleases tedious and time-consuming. Zinc-finger proteins bind to the DNA

1 major groove by three or four "fingers", each of them made of a $\beta\beta\alpha$ structure that coordinates a Zn^{2+}
2 ion and normally contacting three nucleotides at a time (Figure 3B). Lysine side chains of each finger
3 are slightly disordered and have the potential to make multiple hydrogen bonds. This leads to some
4 level of degenerescence in DNA recognition for each finger, exemplified by the binding of GCA
5 triplets by GCT fingers^{78, 79}. TALE repeats bind to the major groove by short α helix-long α helix
6 structures (Figure 3C). Due to their repeated nature, TALEs wrap around DNA, contacting each
7 cognate nucleotide by the RVD loop located between α helices. Finally, the *S. pyogenes* Cas9 nuclease
8 is a much larger protein than its predecessors (1360 aa), and binds its target by an original combination
9 of RNA-DNA hybridization and DNA-protein interactions (Figure 3D). The PAM sequence in the
10 non-target strand is recognized by two conserved arginine residues, resulting in local strand separation
11 immediately upstream of the PAM⁹². This local separation allows hybridization of the guide RNA to its
12 complementary DNA target sequence. Therefore, the presence of a PAM right next to the targeted
13 DNA sequence is essential for Cas9 function. Catalysis is performed by the HNH domain on the
14 complementary DNA strand, and by the RuvC-like domain on the non-complementary strand, leading
15 to the formation of a DSB by two single-strand breaks.
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37

38 **Efficacies and specificities of highly specific nucleases**

39
40 In nematodes, comparable mutation frequencies of the *ben-1* locus were achieved with ZFN (5.1% of
41 progeny) and with TALEN (3.5% of progeny), with a similar mutation spectrum⁹⁴. In zebrafish, a
42 study comparing relative efficacies of 33 active TALENs and ZFNs, concluded to an overall better
43 efficacy of TALENs (20% average somatic mutation rate) as compared to ZFN (5% average mutation
44 rate). However, these results are not totally comparable, since TALENs were assembled with *FokI*
45 homodimers, whereas obligate heterodimers were used with ZFNs⁹⁵. In *Drosophila*, comparisons
46 between TALENs and ZFNs led to the conclusion that architectures containing *FokI* homodimers were
47 at least ten-fold more efficient at generating mutations than architectures using obligate heterodimers.
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 show a higher efficacy, although some TALENs do not work and some ZFNs are equally or more
2 efficient ⁹⁶. Finally, a recent study comparing a CRISPR-Cas9 and a TALEN targeted to the *SORT1*
3 gene in human stem cells, showed a much better efficacy of the Cas9 (76%) compared to the TALEN
4 (11%). Whole-genome sequencing of independent cell clones showed a similar number of off-target
5 indels (6 with TALEN, 10 with Cas9) and a slightly lower, but not statistically significant, number of
6 single-nucleotide mutations for the Cas9 nuclease (74±22 mutations/clone) than for the TALEN
7 (107±46 mutations/clone) ⁹⁷. More similar comparative studies will be needed in the near future,
8 particularly between TALENs and CRISPR-Cas9 nucleases, but it is reasonable to think that both
9 efficacies and specificities will depend widely both on DNA sequences and on genomic regions
10 targeted.
11
12
13
14
15
16
17
18
19
20
21
22
23
24

25 **NUCLEASE DESIGN FOR FUTURE GENE THERAPY**

26
27 Classical gene therapy approaches expressing a trans-complementing gene ⁹⁸, or using oligonucleotide
28 strategies to down-regulate gene expression ⁹⁹, need to be active all life long. In comparison, a
29 nuclease targeted to induce a DSB into a given repeat tract needs only to be expressed transiently.
30 Experiments in yeast showed that a CAG/CTG trinucleotide repeat was shortened below pathological
31 length by a TALEN after three days of continuous expression (Mosbach and Richard, unpublished). In
32 a similar approach, the *I-Sce I* meganuclease was recently transiently expressed in murine T cells
33 mutated for the Artemis gene, and therefore deficient in early B- and T-cell V(D)J recombination. The
34 meganuclease was successfully delivered by an integrase-defective lentiviral vector, and correction of
35 the mutation was achieved by homologous recombination with a wild-type Artemis donor sequence
36 carried by another lentiviral vector ¹⁰⁰.
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51

52 For trinucleotide repeats, once the strategy and the nuclease will be chosen (Box 3), will remain to
53 select a vector to deliver the nuclease into human patients. This part will be challenging, given that for
54 some disorders many different cell types need to be targeted. For example, myotonic dystrophy type 1
55 is a multisystemic disease, characterized by muscle hyperexcitability, progressive myopathy, cataracts,
56
57
58
59
60
61
62
63
64
65

cardiac defects, insulin resistance, testicular atrophy, and various neuropsychiatric manifestations^{7, 101,}
¹⁰². In order to suppress all clinical symptoms, skeletal muscle fibers, smooth muscles, cardiac cells
and neurons will have to be targeted by the nuclease. High-capacity adenoviral vectors lacking all viral
sequences allow gene transfer of up to 16 kb of foreign DNA, sufficient to carry a TALEN or a
CRISPR-Cas9, sequences required for their expression, and stuffer DNA to improve packaging
efficiency. Such vector may be produced in large amounts and may target different cell types,
depending on the serotype used¹⁰³. Recently, CRISPR-Cas9 nucleases directed against the *Pten* or the
p53 genes, were delivered to mice through direct hydrodynamic injection of unpackaged plasmid DNA
via the tail vein. Animals developed liver tumours due to mutations in the targeted genes, showing that
plasmid DNA carrying the nuclease gene could reach target cells and induce mutations¹⁰⁴. However,
tumour induction was quite low, since only 2-3% of cells carried a mutation. This frequency may
probably be improved in the near future by injecting an adenoviral vector carrying the nuclease,
instead of naked plasmid DNA.

REFERENCES

1. Verkerk, A.J.M.H., *et al.* (1991) Identification of a gene (*FMR-1*) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell* 65, 905-914
2. Brook, J.D., *et al.* (1992) Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member. *Cell* 68, 799-808
3. Harley, H.G., *et al.* (1992) Expansion of an unstable DNA region and phenotypic variation in myotonic dystrophy. *Nature* 355, 545-546
4. The Huntington's Disease Collaborative Research Group (1993) A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* 72, 971-983
5. Mandel, J.-L. (1997) Breaking the rule of three. *Nature* 386, 767-769
6. Brown, L.Y., and Brown, S.A. (2004) Alanine tracts: the expanding story of human illness and trinucleotide repeats. *Trends Genet.* 20, 51-58
7. Orr, H.T., and Zoghbi, H.Y. (2007) Trinucleotide repeat disorders. *Annu. Rev. Neurosci.* 30, 575-621
8. Gatchel, J.R., and Zoghbi, H.Y. (2005) Diseases of unstable repeat expansion: mechanisms and common principles. *Nature Reviews Genetics* 6, 743-755
9. O'Hoy, K.L., *et al.* (1993) Reduction in size of the myotonic dystrophy trinucleotide repeat mutation during transmission. *Science* 259, 809-812
10. Cleary, J.D., and Pearson, C.E. (2005) Replication fork dynamics and dynamic mutations: the fork-shift model of repeat instability. *Trends Genet.* 21, 272-280
11. Lenzmeier, B.A., and Freudenreich, C.H. (2003) Trinucleotide repeat instability: a hairpin curve at the crossroads of replication, recombination, and repair. *Cytogenetic and Genome Research* 100, 7-24
12. McMurray, C.T. (2010) Mechanisms of trinucleotide repeat instability during human development. *Nat Rev Genet* 11, 786-799
13. Mirkin, S.M. (2006) DNA structures, repeat expansions and human hereditary disorders. *Curr. Opin. Struct. Biol.* 16, 351-358
14. Pearson, C.E., *et al.* (2005) Repeat instability: mechanisms of dynamic mutations. *Nature Reviews Genetics* 6, 729-742
15. Richard, G.F., *et al.* (2008) Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. *Microbiol. Mol. Biol. Rev.* 72, 686-727

16. Wells, R.D., *et al.* (2005) Advances in mechanisms of genetic instability related to hereditary neurological diseases. *Nucleic Acids Res* 33, 3785-3798
17. Kang, S., *et al.* (1995) Expansion and deletion of CTG repeats from human disease genes are determined by the direction of replication in *E. coli*. *Nature Genet.* 10, 213-217
18. Freudenreich, C.H., *et al.* (1998) Expansion and length-dependent fragility of CTG repeats in yeast. *Science* 279, 853-856
19. Kerrest, A., *et al.* (2009) *SRS2* and *SGS1* prevent chromosomal breaks and stabilize triplet repeats by restraining recombination. *Nature Structural and Molecular Biology* 16, 159-167
20. Shishkin, A.A., *et al.* (2009) Large-scale expansions of Friedreich's ataxia GAA repeats in yeast. *Mol. Cell* 35, 82-92
21. Zhang, Y., *et al.* (2012) Genome-wide screen identifies pathways that govern GAA/TTC repeat fragility and expansions in dividing and nondividing yeast cells. *Mol. Cell* 48, 254-265
22. Cleary, J.D., *et al.* (2010) Tissue- and age-specific DNA replication patterns at the CTG/CAG-expanded human myotonic dystrophy type 1 locus. *Nat Struct Mol Biol* 17, 1079-1087
23. Liu, G., *et al.* (2012) Altered replication in human cells promotes DMPK (CTG)(n) . (CAG)(n) repeat instability. *Mol. Cell. Biol.* 32, 1618-1632
24. Richard, G.-F., *et al.* (2003) Contractions and expansions of CAG/CTG trinucleotide repeats occur during ectopic gene conversion in yeast, by a *MUS81*-independent mechanism. *J. Mol. Biol.* 326, 769-782
25. Richard, G.-F., *et al.* (2000) Recombination-induced CAG trinucleotide repeat expansions in yeast involve the *MRE11/RAD50/XRS2* complex. *EMBO J.* 19, 2381-2390
26. Kovtun, I.V., *et al.* (2007) OGG1 initiates age-dependent CAG trinucleotide expansion in somatic cells. *Nature* 447, 447-452
27. Hubert, L., Jr., *et al.* (2011) Topoisomerase 1 and single-strand break repair modulate transcription-induced CAG repeat contraction in human cells. *Mol. Cell. Biol.* 31, 3105-3112
28. Larrea, A.A., *et al.* (2010) SnapShot: DNA mismatch repair. *Cell* 141, 730 e731
29. Manley, K., *et al.* (1999) *Msh2* deficiency prevents *in vivo* somatic instability of the CAG repeat in Huntington disease transgenic mice. *Nat. Genet.* 23, 471-473
30. Pinto, R.M., *et al.* (2013) Mismatch Repair Genes *Mlh1* and *Mlh3* Modify CAG Instability in Huntington's Disease Mice: Genome-Wide and Candidate Approaches. *PLoS Genet* 9, e1003930
31. Savouret, C., *et al.* (2003) CTG repeat instability and size variation timing in DNA repair-deficient mice. *EMBO J.* 22, 2264-2273
32. Savouret, C., *et al.* (2004) MSH2-dependent germinal CTG repeat expansions are produced continuously in spermatogonia from DM1 transgenic mice. *Mol. Cell. Biol.* 24, 629-637
33. Tome, S., *et al.* (2009) MSH2 ATPase domain mutation affects CTG*CAG repeat instability in transgenic mice. *PLoS Genet* 5, e1000482
34. Tome, S., *et al.* (2013) MSH3 Polymorphisms and Protein Levels Affect CAG Repeat Instability in Huntington's Disease Mice. *PLoS Genet* 9, e1003280
35. Panigrahi, G.B., *et al.* (2010) Isolated short CTG/CAG DNA slip-outs are repaired efficiently by hMutSbeta, but clustered slip-outs are poorly repaired. *Proc. Natl. Acad. Sci. U. S. A.* 107, 12593-12598
36. Panigrahi, G.B., *et al.* (2012) Human mismatch repair protein hMutLalpha is required to repair short slipped-DNAs of trinucleotide repeats. *J. Biol. Chem.* 287, 41844-41850
37. Pluciennik, A., *et al.* (2013) Extrahelical (CAG)/(CTG) triplet repeat elements support proliferating cell nuclear antigen loading and MutLalpha endonuclease activation. *Proc. Natl. Acad. Sci. U. S. A.* 110, 12277-12282
38. Pearson, C.E., *et al.* (1997) Human MSH2 binds to trinucleotide repeat DNA structures associated with neurodegenerative diseases. *Hum. Mol. Genet.* 6, 1117-1123
39. Owen, B.A., *et al.* (2005) (CAG)(n)-hairpin DNA binds to Msh2-Msh3 and changes properties of mismatch recognition. *Nat Struct Mol Biol* 12, 663-670
40. Tian, L., *et al.* (2009) Mismatch recognition protein MutSbeta does not hijack (CAG)n hairpin repair *in vitro*. *J. Biol. Chem.* 284, 20452-20456
41. Richard, G.-F., *et al.* (1999) Double-strand break repair can lead to high frequencies of deletions within short CAG/CTG trinucleotide repeats. *Mol. Gen. Genet.* 261, 871-882
42. Miret, J.J., *et al.* (1997) Instability of CAG and CTG trinucleotide repeats in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 17, 3382-3387
43. Miret, J.J., *et al.* (1998) Orientation-dependent and sequence-specific expansions of CTG/CAG trinucleotide repeats in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* 95, 12438-12443
44. Bhattacharyya, S., and Lahue, R.S. (2005) Srs2 helicase of *Saccharomyces cerevisiae* selectively unwinds triplet repeat DNA. *J. Biol. Chem.* 280, 33311-33317
45. Dhar, A., and Lahue, R.S. (2008) Rapid unwinding of triplet repeat hairpins by Srs2 helicase of *Saccharomyces cerevisiae*. *Nucleic Acids Res* 36, 3366-3373
46. Bhattacharyya, S., and Lahue, R.S. (2004) *Saccharomyces cerevisiae* Srs2 DNA helicase selectively blocks expansions of trinucleotide repeats. *Mol. Cell. Biol.* 24, 7324-7330
47. Anand, R.P., *et al.* (2012) Overcoming natural replication barriers: differential helicase requirements. *Nucleic Acids Res* 40, 1091-1105
48. Frizzell, A., *et al.* (2014) RTEL1 inhibits trinucleotide repeat expansions and fragility. *Cell Rep* 6, 827-835

49. Seigneur, M., *et al.* (1998) RuvAB acts at arrested replication forks. *Cell* 95, 419-430
50. Lopes, M., *et al.* (2001) The DNA replication checkpoint response stabilizes stalled replication forks. *Nature* 412, 557-561
51. Jackson, A., *et al.* (2014) Expansion of CAG repeats in *Escherichia coli* is controlled by single-strand DNA exonucleases of both polarities. *Genetics* 10.1534/Genetics.114.168245
52. Giannattasio, M., *et al.* (2014) Visualization of recombination-mediated damage bypass by template switching. *Nat Struct Mol Biol* 21, 884-892
53. Gacy, A.M., *et al.* (1995) Trinucleotide repeats that expand in human disease form hairpin structures *in vitro*. *Cell* 81, 533-540
54. McMurray, C.T. (1999) DNA secondary structure: a common and causative factor for expansion in human disease. *Proc. Natl. Acad. Sci. USA* 96, 1823-1825
55. Kiliszek, A., and Rypniewski, W. (2014) Structural studies of CNG repeats. *Nucleic Acids Res*
56. Kiliszek, A., *et al.* (2010) Atomic resolution structure of CAG RNA repeats: structural insights and implications for the trinucleotide repeat expansion diseases. *Nucleic Acids Res* 38, 8370-8376
57. Mooers, B.H., *et al.* (2005) The structural basis of myotonic dystrophy from the crystal structure of CUG repeats. *Proc. Natl. Acad. Sci. U. S. A.* 102, 16626-16631
58. Sobczak, K., *et al.* (2003) RNA structure of trinucleotide repeats associated with human neurological diseases. *Nucleic Acids Res* 31, 5469-5482
59. Fry, M., and Loeb, L.A. (1994) The fragile X syndrome d(CGG)_n nucleotide repeats form a stable tetrahelical structure. *Proc. Natl. Acad. Sci. USA* 91, 4950-4954
60. Fojtik, P., and Vorlickova, M. (2001) The fragile X chromosome (GCC) repeat folds into a DNA tetraplex at neutral pH. *Nucleic Acids Res.* 29, 4684-4690
61. Mariappan, S.V., *et al.* (1999) The high-resolution structure of the triplex formed by the GAA/TTC triplet repeat associated with Friedreich's ataxia. *J. Mol. Biol.* 285, 2035-2052
62. Mirkin, S.M., *et al.* (1987) DNA H form requires a homopurine-homopyrimidine mirror repeat. *Nature* 330, 495-497
63. Schaffitzel, C., *et al.* (2001) In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with *Stylonychia lemnae* macronuclei. *Proc. Natl. Acad. Sci. U. S. A.* 98, 8572-8577
64. Axford, M.M., *et al.* (2013) Detection of Slipped-DNAs at the Trinucleotide Repeats of the Myotonic Dystrophy Type I Disease Locus in Patient Tissues. *PLoS Genet* 9, e1003866
65. Samadashwily, G., *et al.* (1997) Trinucleotide repeats affect DNA replication *in vivo*. *Nature Genet.* 17, 298-304
66. Pelletier, R., *et al.* (2003) Replication and expansion of trinucleotide repeats in yeast. *Mol. Cell. Biol.* 23, 1349-1357
67. Kim, H., and Livingston, D.M. (2006) A high mobility group protein binds to long CAG repeat tracts and establishes their chromatin organization in *Saccharomyces cerevisiae*. *J. Biol. Chem.* 281, 15735-15740
68. House, N.C., *et al.* (2014) NuA4 Initiates Dynamic Histone H4 Acetylation to Promote High-Fidelity Sister Chromatid Recombination at Postreplication Gaps. *Mol. Cell* 55, 818-828
69. Gannon, A.M., *et al.* (2012) MutSbeta and histone deacetylase complexes promote expansions of trinucleotide repeats in human cells. *Nucleic Acids Res* 40, 10324-10333
70. Debacker, K., *et al.* (2012) Histone deacetylase complexes promote trinucleotide repeat expansions. *PLoS Biol* 10, e1001257
71. Dion, V., *et al.* (2008) Dnmt1 deficiency promotes CAG repeat expansion in the mouse germline. *Hum. Mol. Genet.* 17, 1306-1317
72. Colleaux, L., *et al.* (1986) Universal code equivalent of a yeast mitochondrial intron reading frame is expressed into *E. Coli* as a specific double strand break endonuclease. *Cell* 44, 521-533
73. Gaj, T., *et al.* (2013) ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol.* 31, 397-405
74. Kim, H., and Kim, J.S. (2014) A guide to genome engineering with programmable nucleases. *Nat Rev Genet* 15, 321-334
75. Arnould, S., *et al.* (2006) Engineering of large numbers of highly specific homing endonucleases that induce recombination on novel DNA targets. *J. Mol. Biol.* 355, 443-458
76. Daboussi, F., *et al.* (2012) Chromosomal context and epigenetic mechanisms control the efficacy of genome editing by rare-cutting designer endonucleases. *Nucleic Acids Res* 40, 6367-6379
77. Kim, Y.G., *et al.* (1996) Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc. Natl. Acad. Sci. U. S. A.* 93, 1156-1160
78. Mittelman, D., *et al.* (2009) Zinc-finger directed double-strand breaks within CAG repeat tracts promote repeat instability in human cells. *Proc. Natl. Acad. Sci. U. S. A.* 106, 9607-9612
79. Liu, G., *et al.* (2010) Replication-dependent instability at (CTG)_x (CAG) repeat hairpins in human cells. *Nat Chem Biol* 6, 652-659
80. Miller, J.C., *et al.* (2007) An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat. Biotechnol.* 25, 778-785
81. Kay, S., *et al.* (2007) A bacterial effector acts as a plant transcription factor and induces a cell size regulator. *Science* 318, 648-651
82. Boch, J., *et al.* (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* 326, 1509-1512

83. Moscou, M.J., and Bogdanove, A.J. (2009) A simple cipher governs DNA recognition by TAL effectors. *Science* 326, 1501
84. Christian, M., *et al.* (2010) Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics* 186, 757-761
85. Richard, G.F., *et al.* (2014) Highly specific contractions of a single CAG/CTG trinucleotide repeat by TALEN in yeast. *PLoS ONE* 9, e95611
86. Boissel, S., *et al.* (2014) megaTALs: a rare-cleaving nuclease architecture for therapeutic genome engineering. *Nucleic Acids Res* 42, 2591-2601
87. Kleinstiver, B.P., *et al.* (2014) The I-TevI nuclease and linker domains contribute to the specificity of monomeric TALENs. *G3 (Bethesda)* 4, 1155-1165
88. Beurdeley, M., *et al.* (2013) Compact designer TALENs for efficient genome engineering. *Nat Commun* 4, 1762
89. Louwen, R., *et al.* (2014) The role of CRISPR-Cas systems in virulence of pathogenic bacteria. *Microbiol. Mol. Biol. Rev.* 78, 74-88
90. Jinek, M., *et al.* (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816-821
91. Doudna, J.A., and Charpentier, E. (2014) Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* 346, 1258096
92. Anders, C., *et al.* (2014) Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* 513, 569-573
93. Sternberg, S.H., *et al.* (2014) DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature*
94. Wood, A.J., *et al.* (2011) Targeted genome editing across species using ZFNs and TALENs. *Science* 333, 307
95. Chen, S., *et al.* (2013) A large-scale in vivo analysis reveals that TALENs are significantly more mutagenic than ZFNs generated using context-dependent assembly. *Nucleic Acids Res* 41, 2769-2778
96. Beumer, K.J., *et al.* (2013) Comparing Zinc Finger Nucleases and Transcription Activator-Like Effector Nucleases for Gene Targeting in *Drosophila*. *G3 (Bethesda)* 3, 1717-1725
97. Veres, A., *et al.* (2014) Low incidence of off-target mutations in individual CRISPR-Cas9 and TALEN targeted human stem cell clones detected by whole-genome sequencing. *Cell Stem Cell* 15, 27-30
98. Fischer, A., *et al.* (2010) 20 years of gene therapy for SCID. *Nat Immunol* 11, 457-460
99. Fiszer, A., and Krzyzosiak, W.J. (2014) Oligonucleotide-based strategies to combat polyglutamine diseases. *Nucleic Acids Res* 42, 6787-6810
100. Riviere, J., *et al.* (2014) Variable correction of Artemis deficiency by I-SceI-meganuclease-assisted homologous recombination in murine hematopoietic stem cells. *Gene Ther.* 21, 529-532
101. Mankodi, A., *et al.* (2000) Myotonic dystrophy in transgenic mice expressing an expanded CUG repeat. *Science* 289, 1769-1772
102. Seznec, H., *et al.* (2001) Mice transgenic for the human myotonic dystrophy region with expanded CTG repeats display muscular and brain abnormalities. *Hum. Mol. Genet.* 10, 2717-2726
103. Jager, L., *et al.* (2009) A rapid protocol for construction and production of high-capacity adenoviral vectors. *Nat Protoc* 4, 547-564
104. Xue, W., *et al.* (2014) CRISPR-mediated direct mutation of cancer genes in the mouse liver. *Nature* 514, 380-384
105. Moure, C.M., *et al.* (2003) The crystal structure of the gene targeting homing endonuclease I-SceI reveals the origins of its target site specificity. *J. Mol. Biol.* 334, 685-695
106. Foster, M.P., *et al.* (1997) Domain packing and dynamics in the DNA complex of the N-terminal zinc fingers of TFIIIA. *Nat. Struct. Biol.* 4, 605-608
107. Deng, D., *et al.* (2012) Structural basis for sequence-specific recognition of DNA by TAL effectors. *Science* 335, 720-723

FIGURE LEGENDS

Figure 1 : Non-pathological and pathological lengths of microsatellites involved in human disorders.

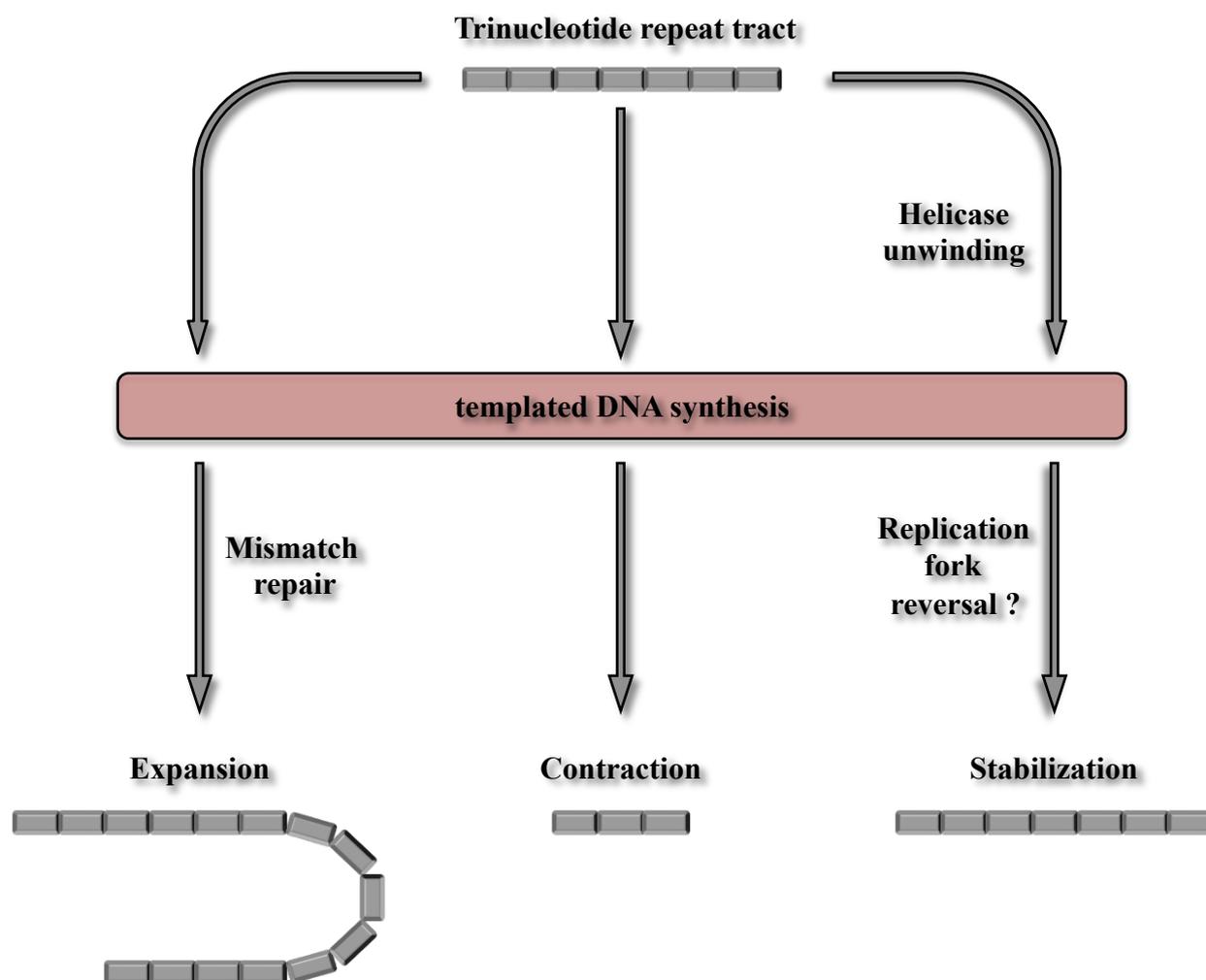
Microsatellites expanded in coding (exons) and non-coding (introns, 3' UTR, 5' UTR) regions are shown on the left. Poly-glutamine disorders include Huntington disease, spinocerebellar ataxia types 1, 2, 3, 6, 7, 17, dentatorubral-pallidoluysian atrophy, spinal and bulbar muscular atrophy. Poly-alanine disorders include synpolydactyly type II, cleidocranial dysplasia, oculopharyngeal muscular dystrophy, holoprosencephaly, hand-foot-genital syndrome and half a dozen other developmental illnesses. On the right, microsatellite sizes in the normal population are shown as blue horizontal bars, whereas expansion sizes found in the affected population are shown as red horizontal bars (log scale). Note that blue and red bars sometimes overlap, since all disorders do not share the same threshold for pathogeny. *: in SCA31, a TGGAA pentanucleotide repeat is found inserted into a more complex (TAGAA)_n (TAAAA TAGAA)_n tandem repeat.

Figure 2 : Trinucleotide repeat contractions and expansions during DSB-repair

A: An I-Sce I recognition site flanked by two short (CAG)₅ repeats was integrated on yeast chromosome X. A homologous copy of the region, containing a CAG/CTG repeat, was available in *trans* as a donor sequence for homologous recombination. **B:** When I-Sce I was induced, yeast cells could repair the double-strand break by intra- or inter-molecular repair. Most of the time (2/3 events), intra-molecular repair was chosen and led to trinucleotide repeat contraction and loss of the I-Sce I site (bottom left). In the other cases, inter-molecular repair by gene conversion led to the loss of the recognition site and expansion or contraction of the repeat tract (bottom right). A representative Southern blot is shown, the dotted line representing the starting size of the repeat tract, blue arrows pointing to repeat contractions, red arrows to repeat expansions, some of them being quite large.

Figure 3 : Crystal structures of the four families of nucleases.

1 For each structure, N and C terminal ends are shown. White bordered inserts show perpendicular view
2 of protein-DNA complexes along DNA axis. **A:** Crystal structure of I-Sce I meganuclease (235 aa)
3 bound to its cognate site. The protein contacts the DNA major groove upstream and downstream the
4 cleavage site, through interactions with sheet 1 ($\beta 1 \beta 2 \beta 3 \beta 4$ structure in red) and sheet 2 ($\beta 5 \beta 6 \beta 7 \beta 8$
5 structure in red). In addition, the N-terminal loop (in red) contacts the DNA minor groove. The two
6 LAGLIDADG motifs are folded into α helices (in yellow). White insert: Detail of catalytic center,
7 showing the three aspartic acid residues involved in coordination of two of the three Ca^{2+} ions (orange
8 balls; PDB structure 1R7M¹⁰⁵). **B:** NMR structure of a three-finger zinc-finger protein, the TFIIIA
9 transcription factor from *Xenopus laevis* (101 aa), bound to 5S RNA gene. Each finger is made of two
10 small β sheets packed against a α helix (in red). White insert: Detail of finger 2, showing the two
11 cysteine and the two histidine residues coordinating a Zn^{2+} ion (orange ball). Note that TFIIIA is a
12 peculiar zinc-finger protein, since instead of contacting 9 base pairs, it shows a more extended contact
13 surface spanning 13 nucleotides (PDB structure 1TF3¹⁰⁶). **C:** Crystal structure of dHax3 TALE (490
14 aa) bound to its cognate site. Each TALE repeat is made of one short and one long α helix, and
15 contacts the DNA major groove by its RVD loop (in red). White insert: Detail of the first TALE repeat
16 encoding the histidine-aspartate RVD contacting its cognate base, cytosine. (PDB structure 3V6T¹⁰⁷).
17
18 **D:** Crystal structure of *Streptococcus pyogenes* Cas9 nuclease (1360 aa) in complex with a single-
19 molecule guide RNA and a target DNA containing a canonical NGG PAM sequence. The HNH
20 nuclease domain is shown in yellow and the RuvC-like domain is shown in light blue. The guide RNA
21 is represented in orange and interacts with its complementary DNA strand (in purple). The two
22 guanines of the PAM (in red) in the non-complementary DNA strand are bound in the major groove by
23 two arginine residues (also in red). White insert: structure of the guide RNA, with the 20 nucleotides
24 hybridized to complementary DNA target shown in red (PDB structure 4UN3⁹²). 3D models were
25 drawn from PDB structures using MacPyMOL.
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

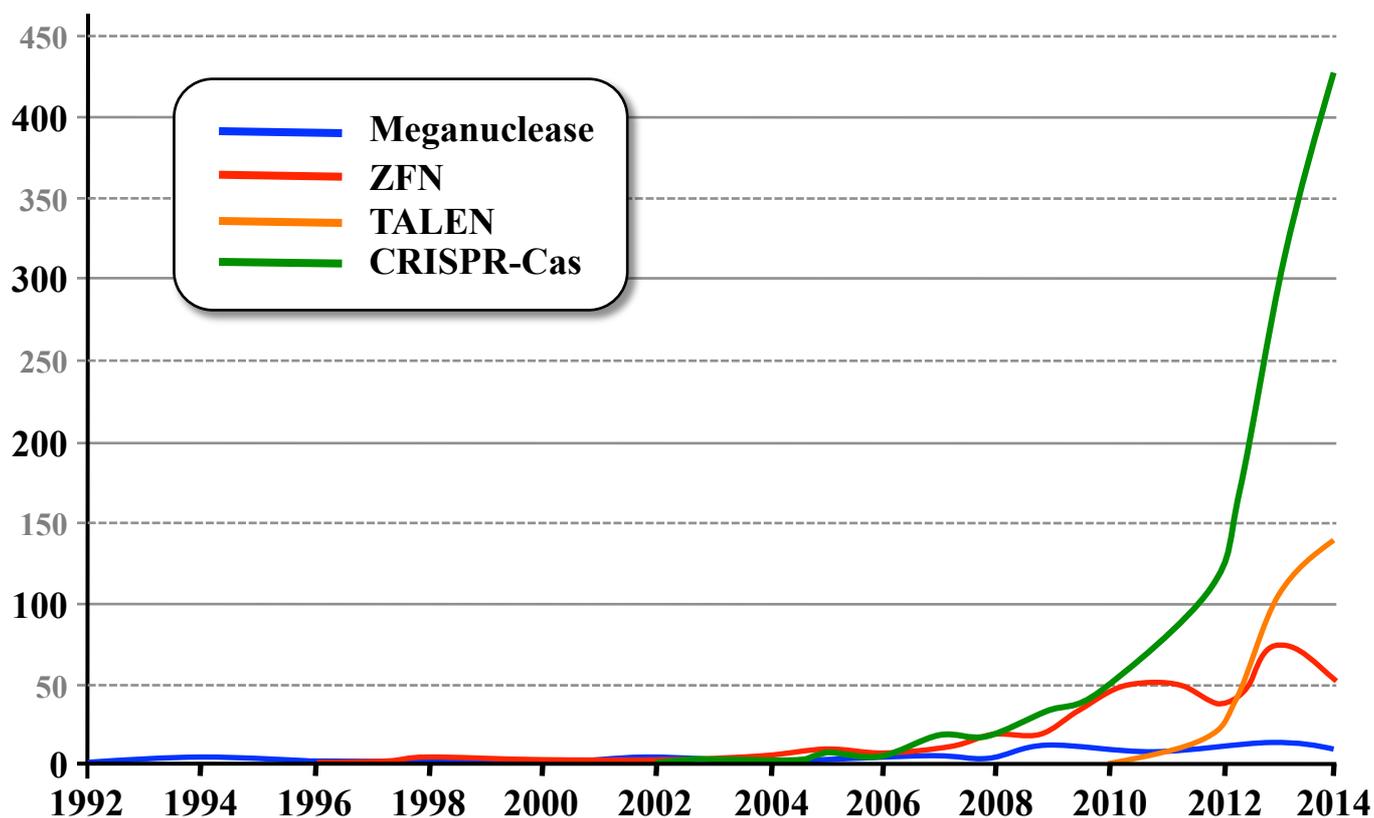


Box 1: Molecular mechanisms leading to trinucleotide repeat instability

Trinucleotide repeat tracts may undergo contractions (length reduction) or expansions (length increase) by any mechanism involving templated synthesis of DNA: replication, double-strand break repair, base excision repair, single-strand break repair, etc. Slippage of the newly synthesized strand on its template occurring during any of these metabolic processes may lead to repeat contraction or expansion, depending whether slippage happens on the newly-synthesized strand or on its template.

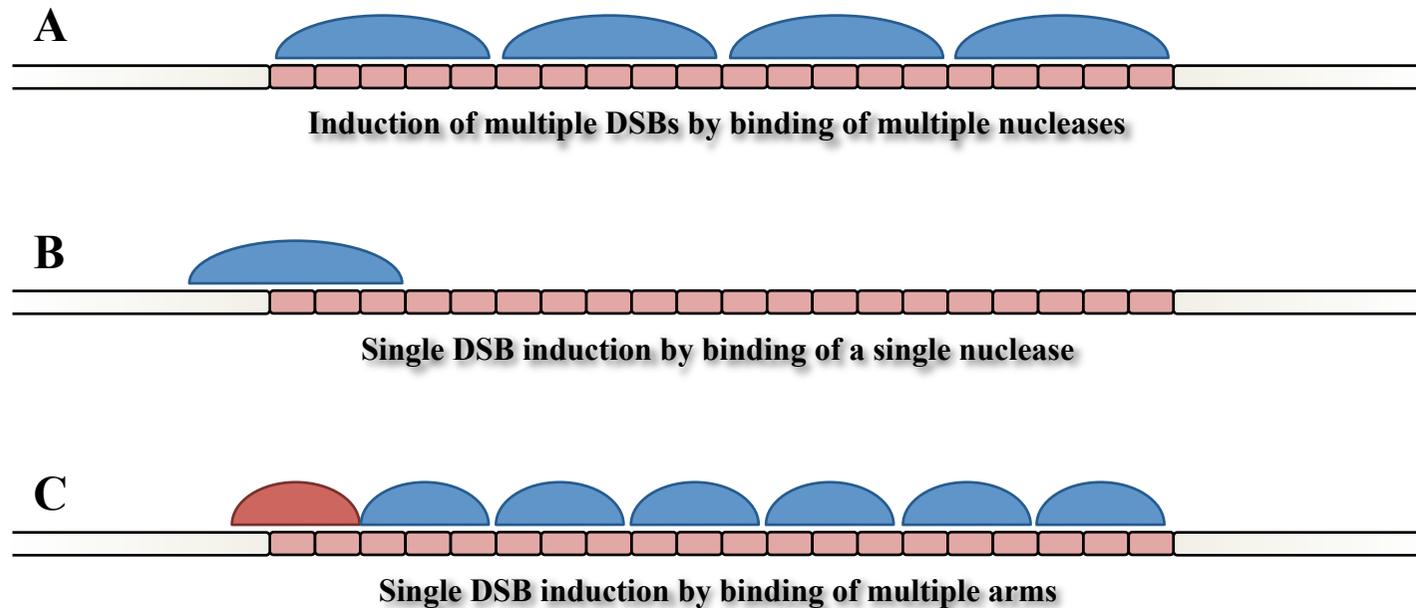
Mismatch repair seems to be specifically involved in trinucleotide repeat expansions, probably by a mechanism involving transient stabilization of a hairpin or slipped-strand DNA structure by Msh2-complexes.

In parallel, secondary structures unwinding by specific DNA helicases (*SGS1*, *SRS2*, *RTEL1*) facilitates DNA synthesis through trinucleotide repeats, reducing slippage, hence increasing stability. Replication fork reversal may be another way of reducing the effect of slippage due to secondary structures. Note the questionmark underlying that direct evidence for replication fork reversal stabilizing trinucleotide repeats is still lacking.



Box 2: Publications mentioning highly-specific nucleases in the PubMed database

Number of hits per year in the PubMed database for each of the following keywords: meganuclease, ZFN, TALEN (or TALE nuclease), CRISPR. The first meganuclease, *I-Sce I*, was discovered in 1986 ⁷², but the term "meganuclease" was coined several years later. The first zinc-finger nuclease was built in 1996 ⁷⁷, and the word appeared in the literature at the same time. TALENs are much more recent and from their engineering in 2009-2010 became very popular, although not as popular as CRISPR-Cas nucleases. Note the dramatic increase of CRISPR articles since 2012, with more than 400 publications in 2014 (as of October 17th).



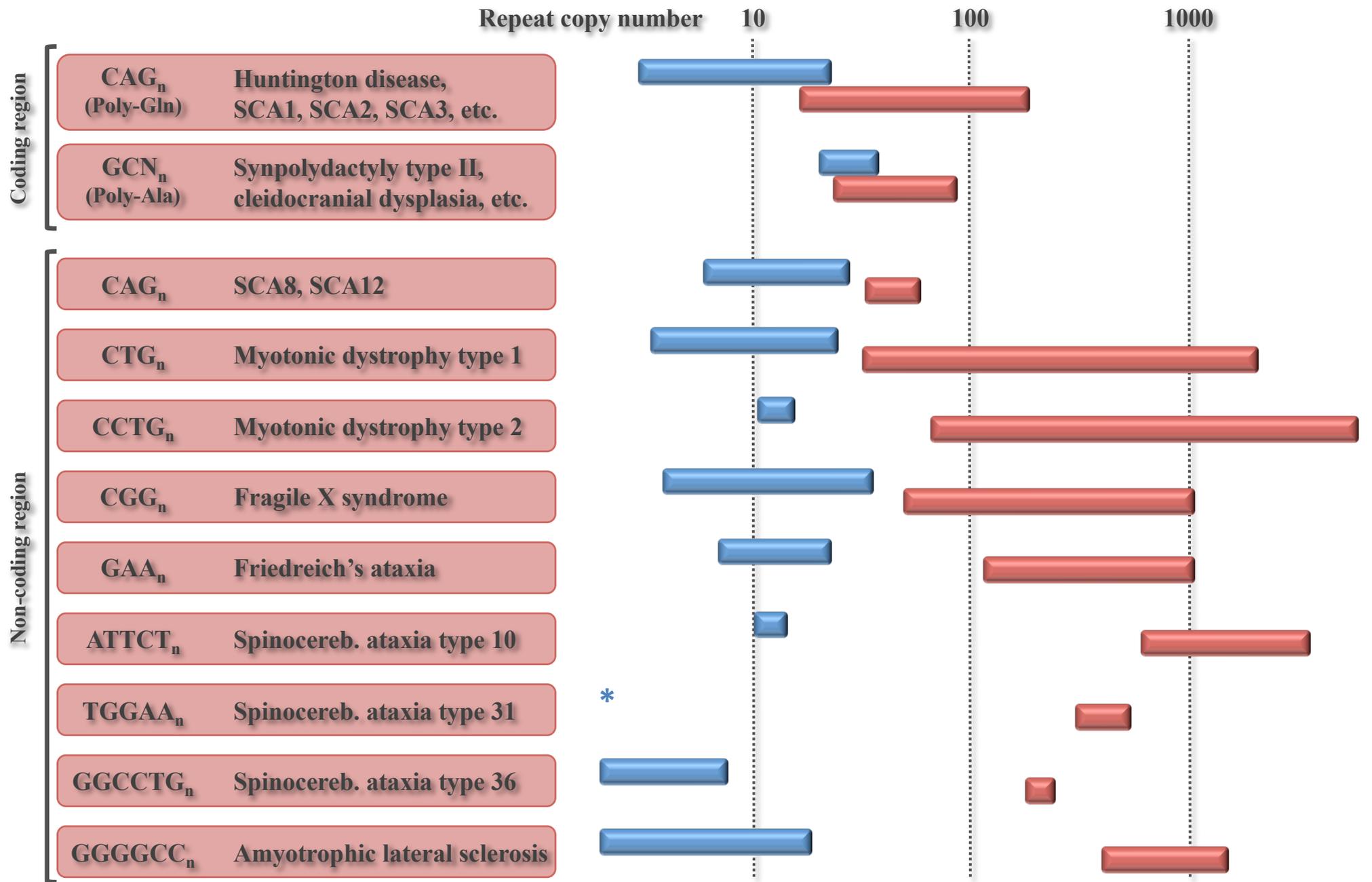
Box 3: Different ways of designing a nuclease targeting a trinucleotide repeat tract

A: The nuclease (in blue) was designed to target the repeated region itself. Several nuclease molecules may bind to the targeted region, leading to the induction of several DSBs within the region. This approach may be used with ZFNs and TALENs, but for CRISPR is restricted to tandem repeats containing a PAM sequence.

B: The nuclease was designed to recognize the junction between repeated and non-repeated DNA. Only one nuclease molecule will be able to bind to this unique sequence. This approach may be used with CRISPR if the junction contains a PAM sequence.

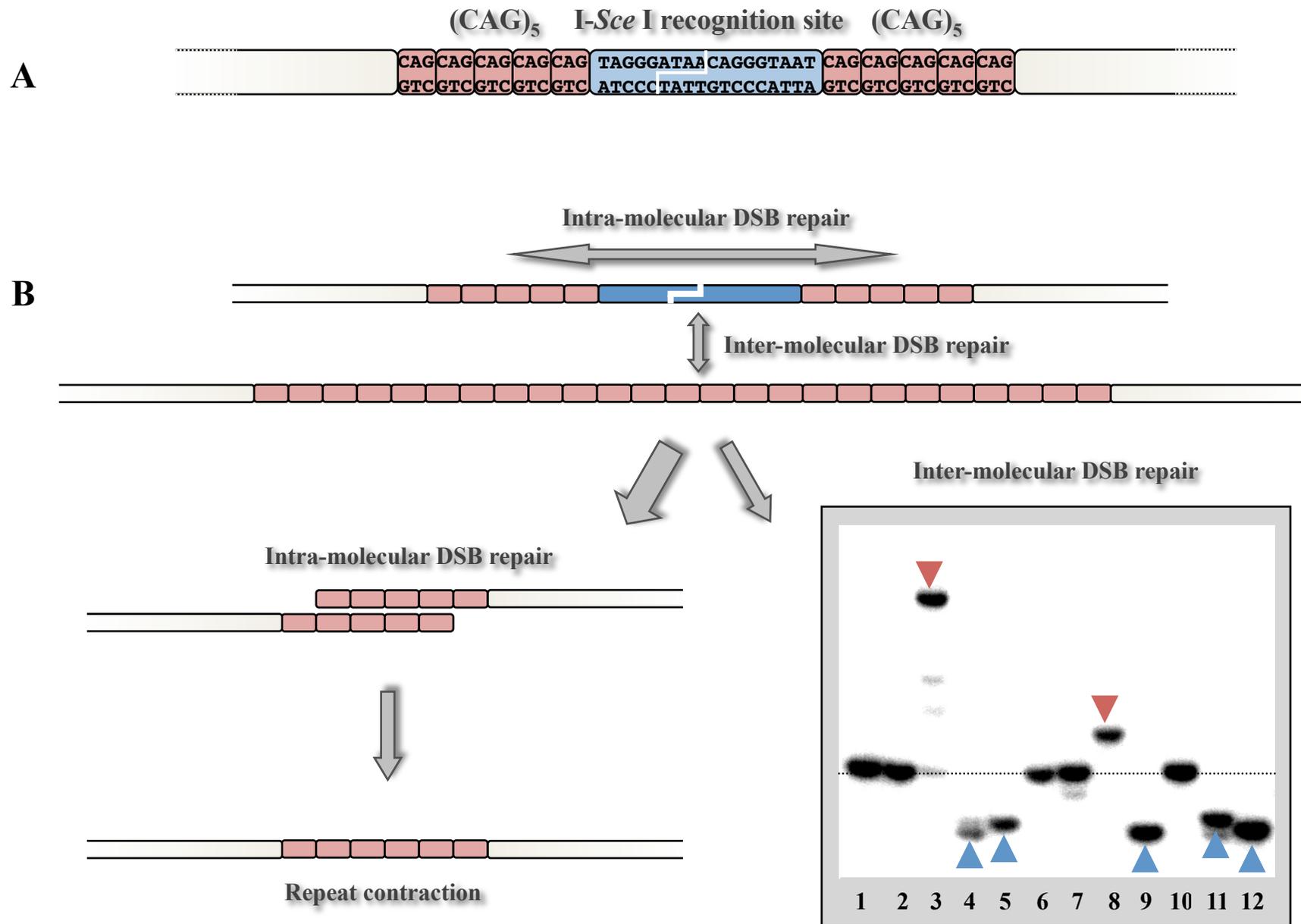
C: The same approach may also be used with ZFNs and TALENs, the left arm binding the junction and the right arm binding the repeat tract. Note that in that case multiple right arms may tandemly bind the repeated region but only one DSB will be induced, at the junction. Whenever possible, this design should be preferred, since off-target binding and cleavage of the nuclease to another repeat tract of the genome will be decreased.

Figure 1



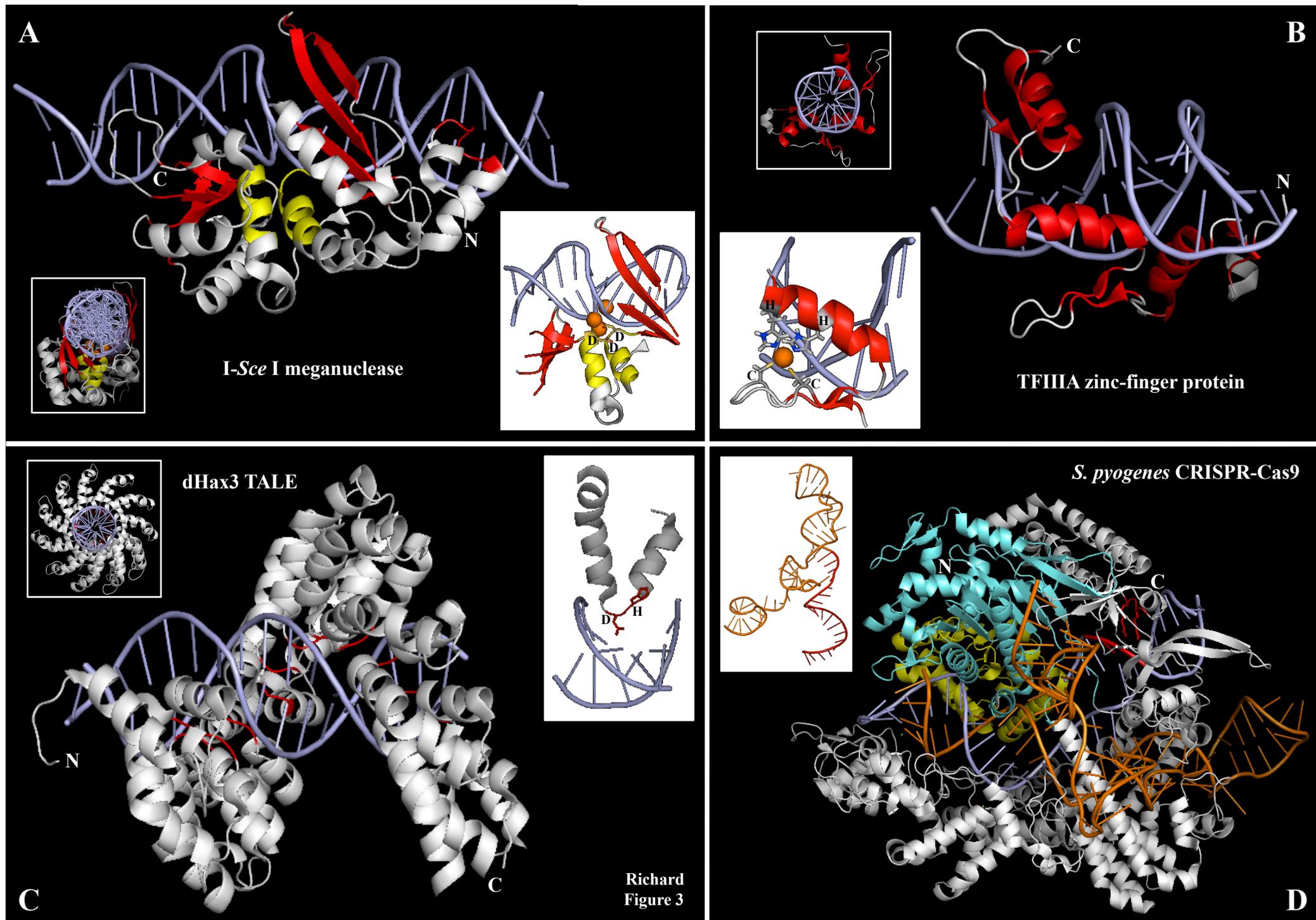
Richard
Figure 1

Figure 2



Richard
Figure 2

Figure 3



Highlight 1

Microsatellite expansions are responsible for more than 30 neurological or developmental disorders

Highlight 2

Gene therapy could involve shortening expanded microsatellites using highly specific nucleases